

# Pivot Machine Translation Using Chinese as Pivot Language

Chao-Hong Liu,<sup>1</sup> Catarina Cruz Silva,<sup>2</sup> Longyue Wang,<sup>1</sup> and Andy Way<sup>1</sup>

<sup>1</sup>ADAPT Centre, Dublin City University, Ireland <sup>2</sup>Unbabel, Portugal

**Abstract.** Pivoting through a popular language with more parallel corpora available (e.g. English and Chinese) is a common approach to build machine translation (MT) systems for low-resource languages. For example, to build a Russian-to-Spanish MT system, we could build one system using the Russian-Spanish corpus directly. We could also build two systems, Russian-to-English and English-to-Spanish, as the resources of the two language pairs are much larger than the Russian-Spanish pair, and use them cascadingly to translate texts in Russian into Spanish by pivoting through English. There are, however, some confusing results on the Pivot MT approach in the literature. In this paper, we reviewed the performance of Pivot MT with the United Nations Parallel Corpus v1.0 (UN6Way) using both English and Chinese as pivot languages. We also report our system performance on the CWMT 2018 Pivot MT shared task, where Japanese patent sentences are translated into English using Chinese as the pivot language.

**Keywords:** Pivot MT · Pivot Language · Patent MT.

## 1 Introduction

The idea of Pivot MT is to build MT systems for a language pair where the availability of its parallel corpus (A–C) is either absent or comparably smaller than the existing parallel corpora paired with a ‘pivot’ language B, i.e. the A–B and B–C corpora [22] [11]. When the availability of parallel corpus A–C is small, taking advantage of A–B and B–C corpora is the main approach to translating sentences from A to C. It is one of the enabling technologies to build MT systems for low-resource languages. There are many strategies in the literature on how to realise this idea in MT systems. Recently it was shown that zero-shot Neural Machine Translation (NMT) could also be trained in the same model for both A-to-C and C-to-A translation directions using only A–B and B–C corpora [6]. However, there is still a big gap on the results compared to the pivot approach of translating with cascading A-to-B and B-to-C models [12].

Two pivot strategies are compared in Utiyama and Isahara (2007), namely phrase-translation and sentence-translation. [22]. In the sentence-translation strategy, the two models (FR-to-EN and EN-to-DE) were used directly. An input French sentence is first translated into an English sentence using the FR-to-EN model and then the MT-ed English sentence is translated into a German sentence using the EN-to-DE model. We refer to this sentence translation strategy as “Naïve Pivot MT” (or “Triangulation” in some literature). In the phrase-translation strategy, two Statistical MT (SMT) models are trained (FR-to-EN and EN-to-DE) and the phrase translation probabilities from the two

phrase-tables are used to create a FR-to-DE phrase-table, which is then used along with a monolingual German language model (LM) in the FR-to-DE MT system.

In Wu and Wang (2007), translation probabilities are interpolated using a small bilingual corpus. The method calculates phrase-translation probabilities and lexical weights from Source-to-Pivot and Pivot-to-Target MT models. The interpolated model for SMT [24] increased BLEU score by one point using 22,000 pairs of Chinese–Japanese parallel data [15].

The zero-shot translation approach, where only one neural network is trained with corpora of several translation pairs and directions, has also been proposed [6]. For example, in the training of that single neural network, Portuguese-to-English and English-to-Spanish directions are both used, with the idea that the one network is able to translate from Portuguese to Spanish, even though there is no direct Portuguese-to-Spanish parallel data used in training. However, in a later review of the approach, the scores using the UN6Way corpus [27] for Pivot MT are below 10 in terms of BLEU in most translation directions [12].

In this paper, we examine the idea of Pivot MT using the Naïve Pivot MT approach for comparison purposes. Both SMT and NMT approaches are employed as base models in the experiments. Our goal is to give an overview of the performance of Pivot MT in a fair setting and to clarify some confusing results reported in the literature, e.g. pivoting through English performed better than models trained with direct parallel corpora using the JRC-Acquis corpus [20,8].

The rest of the paper is organised as follows. In Section 2, we give an introduction to Pivot Machine Translation. In Section 3, our experiments are presented, followed by discussion in Section 4. Conclusions are given in Section 5.

## 2 Pivot Machine Translation

Pivot MT is the technology that we use to build A-to-C and/or C-to-A MT systems without (or with little) parallel data of the A–C language pair. A ‘pivot’ language B could be used to help build A–C MT systems if there are decent sizes of A–B and B–C parallel corpora to be taken advantage of [22,24,8,6].

In addition to the main Pivot MT approaches mentioned in Section 1, there are several strategies proposed to further improve pivoting performance. A joint training algorithm is introduced to connect the two separate models in the training phase [2]. Further work on the use of word embeddings in the pivot language is also suggested for Pivot NMT systems [6]. A method incorporating Markov random walks is introduced to alleviate the error propagation problem in Pivot MT, by connecting translation phrases of source and target languages [26].

A Teacher-Student Framework for zero resource NMT is proposed in [1]. The idea is to use a pivot-to-target NMT model (as “teacher”) to guide the training of the source-to-target (the “student”) model, in which source–target parallel data is not available. The framework might also work using SMT systems, but no experimentation exists on this.

An NMT-based pivot translation method has been proposed [5]. The architecture used in its ‘one-to-one’ strategy is the same as the “sentence translation strategy” described in [22]. The only difference is that SMT models are replaced by NMT models.

A single attention model is introduced to be shared across all language pairs, which enables the training of multi-way translation system in one NMT model [5]. Accordingly, the second strategy proposed in [5] is the use of ‘many-to-one’ translation in pivot MT. The strategy is while translating from ES to FR, the Spanish sentence is first translated into English using the ES-to-EN NMT model, and then from ‘both’ the original Spanish sentence and the MT-ed English sentence, into a French sentence using a “multi-way multilingual NMT” model. However, the two strategies do not perform well in the reports [5].

### 3 Experiments

We conduct our experiments on both SMT and NMT models. We used the case-insensitive 4-gram BLEU metric [15] for evaluation, and *sign-test* [3] for statistical significance testing.

We employ Moses [9] to build our phrase-based SMT models. The 5-gram language models are trained using the SRI Language Toolkit [21]. To obtain word alignment, we run GIZA++ [14] on the training data together with News-Commentary11 corpora. We use minimum error rate training [13] to optimize the feature weights. The maximum length of sentences is set as 80.

We employ an attentional encoder-decoder architecture as described in [16] using the Marian framework<sup>1</sup> [7], implemented in C++. We pre-process the data with similar routines in Moses<sup>2</sup> [9], using the following steps: entity replacement (applied to numbers, emails, urls and alphanumeric entities), tokenization, truecasing and byte-pair encoding (BPE) [17] with 89,500 merge operations. The models are trained on sentences of lengths up to 50 words with early stopping. Mini-batches were shuffled during processing with a mini-batch size of 80 sentences. The word-embedding dimension and the hidden layer size are 512. We selected the model that yields the best performance on the validation set.

For the experiments using the UN corpus, we built three MT systems (A-to-B, B-to-C and A-to-C) for each pivot triplet (A-B-C). The base MT model is either SMT or NMT. We used the default settings of Moses 4.0 as the base SMT model, and the Transformer model as implemented in [25] as the base NMT model.

There are more than ten million sentence pairs in the UN6Way corpus [4]. In addition to using the complete set of sentence pairs, we also randomly chose 500K sentence pairs for the experiments. This random subset of UN6Way Corpus is referred to as UN6Way-500K in this paper in order to investigate the effect of increased training data size. The corpus contains the same sentences in each of the six languages, i.e. Arabic, Chinese, English, French, Russian and Spanish. However, we do not include experiments involving Arabic (in both SMT and NMT systems) and Russian (in SMT systems) as they require additional pre-processing and post-processing.

Chinese sentences are segmented using the open-source Jieba segmenter<sup>3</sup> [23]. Segmented Chinese sentences are used as source and target for the MT system training

<sup>1</sup> <https://marian-nmt.github.io/>

<sup>2</sup> <http://www.statmt.org/moses/>

<sup>3</sup> <https://github.com/fxsjy/jieba>

and test data. No additional pre-processing and post-processing tools are used. Likewise, tokenised English, French and Spanish following Moses 4.0 default settings are used as source and target for training and test data. Our experiments focus on comparing the MT performance with and without pivoting, i.e. A-to-C versus A-to-B-to-C using B as pivot.

### 3.1 Results of Direct MT Systems

The performance of SMT systems trained with the UN6Way-500K corpus is shown in Table 1. The results are obtained using direct (i.e. A-to-C) MT systems. We can see from the table that the BLEU scores of translations to and from Chinese are much lower than translations between any two of the three European languages (English, French and Spanish).

Looking at the scores of the two translation directions of one language pair in Table 1, it can be seen that inter-translations between two of the three languages, English, French and Spanish, are of the same MT performance in terms of BLEU scores. For example, EN-to-ES and ES-to-EN are 47.77 and 46.45, respectively. For translation pairs involving Chinese and Russian, however, the performance is quite different between the two translation directions of a language pair. For example, ZH-to-ES is 31.14 in terms of BLEU and ES-to-ZH is 18.91. There are more than 10 points difference in general between translations to and from Chinese.

**Table 1:** Evaluation of baseline Statistical Machine Translation (SMT) systems using 500K pairs of UN6Way corpus to simulate a low-resource scenario

BLEU [%]	Target				
	EN	ZH	RU	ES	FR
EN		22.49	30.59	47.77	41.57
ZH	32.20		20.81	31.14	28.22
RU	40.23	19.38		39.32	35.24
ES	46.45	18.91	27.76		40.40
FR	41.80	17.80	26.66	43.75	

The performance of direct NMT systems trained with the UN6Way-500K corpus is shown in Table 2. We can also observe that scores of translations to and from Chinese are lower. However, NMT systems in general performed better than SMT systems to and from Chinese. Using the UN6Way-500K corpus for MT training, SMT performed better in some translation pairs and directions, e.g. FR-to-EN and ES-to-RU, and NMT performed better in others, e.g. ZH-to-EN and FR-to-ZH.

The results also show that despite UN6Way-500K being a relatively small corpus for NMT training, NMT models are able to outperform their SMT counterparts in most language pairs and translation directions involving Chinese. We believe this is because SMT relies on word segmenters to pre-process Chinese sentences, while NMT systems incorporate BPE to learn subword units during the training [18]. For other language

pairs and translation directions, however, SMT outperformed NMT trained with small corpora.

**Table 2:** Evaluation of baseline Neural Machine Translation (NMT) systems using 500K pairs of UN6Way corpus to simulate a low-resource scenario

BLEU [%]	Target				
	EN	ZH	RU	ES	FR
EN		30.94	29.66	41.75	34.92
ZH	32.88		22.09	28.27	25.17
RU	36.82	25.22		31.70	28.40
ES	41.20	24.79	25.50		35.07
FR	37.12	23.10	23.51	37.16	

The performance of SMT and NMT systems trained with the whole UN6Way corpus is shown in Table 3 and Table 4, respectively. We can still observe that translations to and from Chinese are lower in general, but the differences between those language pairs not involving Chinese are smaller.

For direct SMT systems, when the size of the training corpus is increased from 500K to 11M, the BLEU scores improve by 10 points in general. Systems translating into Chinese were observed to have a bigger improvement compared to other language pairs and translation directions, e.g. English-to-Chinese improves from 22.49 to 37.87 in terms of BLEU.

**Table 3:** Evaluation of base SMT systems using the complete UN6Way corpus (11M pairs)

BLEU [%]	Target				
	EN	ZH	RU	ES	FR
EN		37.87	43.29	61.22	50.07
ZH	42.88		29.61	39.65	34.49
RU	52.62	32.60		49.58	43.31
ES	59.83	31.25	39.72		49.70
FR	52.20	30.05	36.53	52.40	

### 3.2 Results of Pivot MT Systems

In this section, the results of our Pivot MT systems are shown. They are derived from the same base systems in Tables 1 and 2. The scores of \*-direct systems are repeated from either Table 1, 2 or 4, for easier comparison with results using Pivot MT.

Table 5 shows the results of pivoting through English using SMT base systems trained with the UN6Way-500K corpus. It shows that for French and Spanish, direct

**Table 4:** Evaluation of base NMT systems using the complete UN6Way corpus (11M pairs)

BLEU [%]	Target				
	EN	ZH	RU	ES	FR
EN		42.64	43.72	52.74	47.19
ZH	47.72		38.00	41.79	36.76
RU	48.39	35.46		41.67	38.23
ES	56.95	37.87	41.02		45.55
FR	48.28	34.03	36.58	46.13	

MT in general outperformed pivoting through English by one to two points in terms of BLEU.

**Table 5:** Evaluation of SMT systems using EN as pivot language with the 500K sample of data

BLEU [%]	Target			
	ZH	RU	ES	FR
ZH-en-pivot		19.81	30.68	27.52
RU-en-pivot	18.62		37.87	33.93
ES-en-pivot	19.30	27.23		38.47
FR-en-pivot	18.54	25.57	40.61	
ZH-direct		20.81	31.14	28.22
RU-direct	19.38		39.32	35.24
ES-direct	18.91	27.76		40.40
FR-direct	17.80	26.66	43.75	

Table 6 shows the results of pivoting through English using NMT base systems. It shows pretty much the same comparative results as those using SMT. For French and Spanish, the performance of pivoting through English is lower than direct NMT by two BLEU points. For translation directions involving Chinese, the performance is comparable. In general, comparing Tables 5 and 6, we see that performance with NMT is 2–5 BLEU points better than SMT. However, for some language pairs and translation directions (e.g. RU-to-ES), the SMT performance is much better (almost 8 BLEU points) than that of NMT. This is also observed in results using the complete set as training data. This experimental result will be examined further in future work.

Table 7 shows the results of pivoting through English using NMT base systems where the whole UN6Ways corpus is used for training. The impact of using more data is significant. By increasing the training from 500K to 11M, the BLEU scores have increased by 10 points in general for both direct models and pivot models using English as pivot language. The gaps between results of direct models and pivot models are larger. This indicates that the pivot strategy is more suitable to be used in small corpus, and this is the situation we would like to employ it.

**Table 6:** Evaluation of NMT systems using EN as pivot language with the 500K sample of data

BLEU [%]	Target			
	ZH	RU	ES	FR
ZH-en-pivot		20.47	27.91	24.89
RU-en-pivot	23.70		31.49	28.15
ES-en-pivot	24.31	24.62		31.32
FR-en-pivot	23.11	22.96	33.29	
ZH-direct		22.09	28.27	25.17
RU-direct	25.22		31.70	28.40
ES-direct	24.79	25.50		35.07
FR-direct	23.10	23.51	37.16	

**Table 7:** Evaluation of NMT systems using EN as pivot language with the complete UN6Way corpus (11M pairs)

BLEU [%]	Target			
	ZH	RU	ES	FR
ZH-en-pivot		33.76	40.41	36.54
RU-en-pivot	35.06		41.74	38.14
ES-en-pivot	36.73	37.70		41.96
FR-en-pivot	33.46	34.48	42.77	
ZH-direct		38.00	41.79	36.76
RU-direct	35.46		41.67	36.76
ES-direct	37.87	41.02		45.55
FR-direct	34.03	36.58	46.13	

### 3.3 Impact of Pivot Choice

In addition to using English as pivot, we also conduct experiments using Chinese as the pivot language. Table 8 shows the results of pivoting through Chinese using SMT base systems trained with the UN6Way-500K corpus. One notable result is that the MT performance pivoting through Chinese to and from English, French and Spanish, is much lower than direct MT models by twelve BLEU points on average. The results are intuitive and confirm that it is beneficial to choose a pivot language that is linguistically close to both source and target languages.

Table 9 shows the results of pivoting through Chinese using NMT base systems. It shows similar comparative results to those using SMT in Table 8. The gains replacing SMT base models with NMT ones are smaller (one to two points improvement in BLEU) compared to those using English as pivot language (four points improvement).

**Table 8:** Evaluation of SMT systems using ZH as pivot language with 500K sample

BLEU [%]	Target			
	EN	RU	ES	FR
EN-zh-pivot		23.94	34.64	30.87
RU-zh-pivot	29.06		29.49	26.72
ES-zh-pivot	31.50	21.48		29.34
FR-zh-pivot	29.83	21.05	31.12	
RU-en-pivot			37.87	33.93
ES-en-pivot		27.23		38.47
FR-en-pivot		25.57	40.61	

**Table 9:** Evaluation of NMT systems using ZH as pivot language with 500K sample

BLEU [%]	Target			
	EN	RU	ES	FR
EN-zh-pivot		20.92	27.10	24.05
RU-zh-pivot	25.54		23.86	21.70
ES-zh-pivot	26.08	18.52		22.54
FR-zh-pivot	24.17	17.89	23.86	
RU-en-pivot			31.49	28.15
ES-en-pivot		24.62		31.32
FR-en-pivot		22.96	33.29	

### 3.4 Results of Japanese-to-English MT Using Chinese as Pivot Language

We participated in the CWMT 2018 shared task on Pivot MT. In this shared task, training corpora are given for the Japanese–Chinese and Chinese–English pairs in the patent domain. Participants trained the systems to translate from Japanese sentences into English using Chinese as the pivot language. We followed the same experimental setup as used for the UN6Way experiments, except pre-processing the segmentations on the Japanese and Chinese corpora. Common sequences of characters that appear in both Japanese and Chinese corpora are extracted (as parallel texts) from the training corpus and they are treated as ‘words’ by longest-word-first segmenters which were used to segment both Japanese and Chinese training corpora. The results of our system (designated as ‘je-2018-S1-primary-a’) is shown in Table 11. Our system took 4th place (out of 5) according to BLEU4-SBP score, but first place in terms of METEOR [10] and Translation Edit Rate (TER) [19].

## 4 Discussions

Our experiments using both SMT and NMT showed that pivoting will lose around 4 points compared to training with direct parallel data of comparable sizes. In [8], pivoting



**Table 10:** Evaluation of NMT systems using ZH as pivot language with the complete UN6Way corpus (11M pairs)

BLEU [%]	Target			
	EN	RU	ES	FR
EN-zh-pivot		34.60	40.66	35.83
RU-zh-pivot	39.21		36.42	32.84
ES-zh-pivot	40.37	31.73		34.44
FR-zh-pivot	36.51	29.62	36.10	
RU-en-pivot			41.74	38.14
ES-en-pivot		37.70		41.96
FR-en-pivot		34.48	42.77	

**Table 11:** Results of Pivot MT (Japanese-to-English) systems using Chinese as pivot language

Systems	BLEU4-SBP	NIST5	METEOR	TER
je-2018-S18-primary-a	0.4124	8.8276	0.3139	0.5297
je-2018-S20-primary-a	0.3904	8.6592	0.3075	0.5416
je-2018-S22-primary-a	0.3656	8.4550	0.2905	0.5636
<b>je-2018-S1-primary-a</b>	0.3428	8.2311	<b>0.3525</b>	<b>0.4811</b>
je-2018-S24-primary-a	0.3410	8.0863	0.3442	0.4926

through English actually performed better than training MT in the direct language pair, in the JRC-Acquis corpus in the legal domain [20]. This finding is now not observed in our experiments using UN6Way.

For this result reported in [8], one possible cause might be that the corpus is curated ‘aligned’ around English, which might give pivoting through English an advantage compared to direct MT training on that particular corpus. Another reason might be that many texts in the JRC-Acquis corpus are in English in their original form [20]. Texts in the other languages are likely to be translations of their English counterparts. This would also give English an advantage when it is the pivot and explain why it performs better in pivot scenarios using the JRC-Acquis corpus.

## 5 Conclusions

In this paper we have reviewed major approaches to Pivot MT. Experiments using Naïve Pivot MT approaches were conducted to review the applicability of Pivot MT systems. Firstly, there were claims stating that pivoting through English outperformed direct trained MT systems. We found that using both the whole UN6Way Corpus and its random subset of 500K sentences pairs, direct MT systems still outperform Pivot MT systems in general. Even when a very different language (i.e. Chinese to-or-from English, French and Spanish) is involved, their performance is still comparable. Secondly, the results showed in general that it would be much more beneficial to choose a pivot language that

is linguistically close to the source and target languages. Thirdly, the results confirm that the errors introduced by pivoting do propagate to the target language. Therefore, it might be necessary to incorporate quality estimation and/or automatic/human post-editing to the intermediate translation of the pivot language, in application scenarios where high-quality translations are demanded.

## Acknowledgements

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant No. 13/RC/2106) and is co-funded under the European Regional Development Fund. This work has partially received funding from the European Union’s Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie Actions (Grant No. 734211; the EU INTERACT project). The project aimed at researching translation in crisis scenarios. Work Package 4 (WP4) of INTERACT project focuses on developing and evaluating Pivot MT engines for specific language pairs including Arabic, Greek and Swahili.

## References

1. Chen, Y., Liu, Y., Cheng, Y., Li, V.O.: A teacher-student framework for zero-resource neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 1925–1935 (2017)
2. Cheng, Y., Yang, Q., Liu, Y., Sun, M., Xu, W.: Joint training for pivot-based neural machine translation. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). pp. 3974–3980. Melbourne, Australia (2017)
3. Collins, M., Koehn, P., Kucerova, I.: Clause restructuring for statistical machine translation. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics. pp. 531–540. Ann Arbor, Michigan, USA (2005)
4. Eisele, A., Chen, Y.: Multium: A multilingual corpus from united nation documents. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010). pp. 2868–2872. Malta (2010)
5. Firat, O., Cho, K., Sankaran, B., Vural, F.T.Y., Bengio, Y.: Multi-way, multilingual neural machine translation. *Computer Speech & Language* **45**, 236–252 (2017)
6. Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., Dean, J.: Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics* **5**, 339–351 (2017)
7. Junczys-Dowmunt, M., Dwojak, T., Hoang, H.: Is neural machine translation ready for deployment? a case study on 30 translation directions. In: Proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT). pp. 1–8. Seattle, WA (2016)
8. Koehn, P., Birch, A., Steinberger, R.: 462 machine translation systems for europe. In: Proceedings of the Twelfth Machine Translation Summit. pp. 65–72. Denver, Colorado, USA (2009)
9. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E.: Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. pp. 177–180. Prague, Czech Republic (2007)

10. Lavie, A., Agarwal, A.: Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments. In: Proceedings of the Second Workshop on Statistical Machine Translation. pp. 228–231. StatMT '07, Prague, Czech Republic (2007)
11. Liu, S., Wang, L., Liu, C.H.: Chinese-portuguese machine translation: A study on building parallel corpora from comparable texts. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). pp. 1485–1494. Miyazaki, Japan (2018)
12. Miura, A., Neubig, G., Sudoh, K., Nakamura, S.: Tree as a pivot: Syntactic matching methods in pivot translation. In: Proceedings of the Second Conference on Machine Translation, Volume 1: Research Paper. pp. 90–98. Copenhagen, Denmark (2017)
13. Och, F.J.: Minimum error rate training in statistical machine translation. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics. pp. 160–167. Sapporo, Japan (2003)
14. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. *Computational Linguistics* **29**(1), 19–51 (2003)
15. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting on association for computational linguistics. pp. 311–318. Philadelphia, PA, USA (2002)
16. Sennrich, R., Firat, O., Cho, K., Birch, A., Haddow, B., Hirschler, J., Junczys-Dowmunt, M., Läubli, S., Miceli Barone, A.V., Mokry, J., Nadejde, M.: Nematus: a toolkit for neural machine translation. In: Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics. pp. 65–68. Valencia, Spain (2017)
17. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 1715–1725. Berlin, Germany (2016)
18. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7–12, 2016, Berlin, Germany, Volume 1: Long Papers. pp. 1715–1725 (2016)
19. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J.: A study of translation edit rate with targeted human annotation. In: Proceedings of the 7th Biennial Conference of the Association for Machine Translation in the Americas (AMTA-2006). pp. 223–231. Cambridge, Massachusetts, USA (2006)
20. Steinberger, R., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufis, D., Varga, D.: The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC-2006). pp. 2142–2147. Genoa, Italy (2006)
21. Stolcke, A.: Srilm - an extensible language modeling toolkit. In: Proceedings of the 7th International Conference on Spoken Language Processing. pp. 901–904. Colorado, USA (2002)
22. Utiyama, M., Isahara, H.: A comparison of pivot methods for phrase-based statistical machine translation. In: Proceedings of Human Language Technologies, The Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2007). pp. 484–491. Rochester, USA (2007)
23. Wang, M.H., Lei, C.L.: Boosting election prediction accuracy by crowd wisdom on social forums. In: Consumer Communications & Networking Conference (CCNC), 2016 13th IEEE Annual. pp. 348–353. IEEE, Las Vegas, USA (2016)
24. Wu, H., Wang, H.: Pivot language approach for phrase-based statistical machine translation. *Machine Translation* **21**(3), 165–181 (2007)

25. Zhang, J., Ding, Y., Shen, S., Cheng, Y., Sun, M., Luan, H., Liu, Y.: Thumt: An open source toolkit for neural machine translation. arXiv preprint arXiv:1706.06415 (2017)
26. Zhu, X., He, Z., Wu, H., Wang, H., Zhu, C., Zhao, T.: Improving pivot-based statistical machine translation using random walk. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 524–534. Seattle, USA (2013)
27. Ziemski, M., Junczys-Dowmunt, M., Pouliquen, B.: The united nations parallel corpus v1.0. In: Proceedings of The International Conference on Language Resources and Evaluation (LREC). pp. 1–5. Portorož, Slovenia (2016)